International Journal of Modeling, Simulation, and Scientific Computing
Vol. 12, No. 2 (2021) 2150012 (18 pages)

DOI: 10.1142/S1793962321500124

Prediction of diabetes and hypertension using multi-layer perceptron neural networks

Hani Bani-Salameh[*,§], Shadi M. Alkhatib[*], Moawyiah Abdalla[*], Mo'taz Al-Hami[†], Ruaa Banat[*], Hala Zyod[*] and Ahed J Alkhatib[‡]

[*]*Department of Software Engineering*
The Hashemite University, Zarqa, Jordan

[†]*Department of Computer Information System*
The Hashemite University, Zarqa, Jordan

[‡]*Department of Legal Medicine*
School of Medicine, JUST, Jordan
[§]*hani@hu.edu.jo*

Received 16 December 2019
Accepted 20 October 2020
Published 4 January 2021

Background: Diabetes and hypertension are two of the commonest diseases in the world. As they unfavorably affect people of different age groups, they have become a cause of concern and must be predicted and diagnosed well in advance.

Objective: This research aims to determine the effectiveness of artificial neural networks (ANNs) in predicting diabetes and blood pressure diseases and to point out the factors which have a high impact on these diseases.

Sample: This work used two online datasets which consist of data collected from 768 individuals. We applied neural network algorithms to predict if the individuals have those two diseases based on some factors. Diabetes prediction is based on five factors: *age, weight, fat-ratio, glucose, and insulin*, while blood pressure prediction is based on six factors: *age, weight, fat-ratio, blood pressure, alcohol, and smoking.*

Method: A model based on the Multi-Layer Perceptron Neural Network (MLP) was implemented. The inputs of the network were the factors for each disease, while the output was the prediction of the disease's occurrence. The model performance was compared with other classifiers such as *Support Vector Machine* (*SVM*) and *K-Nearest Neighbors* (*KNN*). We used performance metrics measures to assess the accuracy and performance of MLP. Also, a tool was implemented to help diagnose the diseases and to understand the results.

Result: The model predicted the two diseases with correct classification rate (CCR) of 77.6% for diabetes and 68.7% for hypertension. The results indicate that MLP correctly predicts the probability of being diseased or not, and the performance can be significantly increased compared with both SVM and KNN. This shows MLPs effectiveness in early disease prediction.

Keywords: Artificial Neural Network (ANN); Multi-Layer Perceptron (MLP); SVM; KNN; decision-making; prediction tools; diabetes; blood pressure; hypertension; software tools.

[§]Corresponding author.

1. Introduction

Diabetes is a result of high blood sugar over a long period. This is caused by defects in the secretion of insulin, insulin work, genetic factors (family history). The disease is one of the commonest diseases in the world. Many studies indicate that more than 80% of people die from diseases caused by diabetes..[1]

High blood pressure (hypertension) disease is widely spread among people. It is important to diagnose and control it as early as possible. Hypertension is considered as one of the main risk factors for heart disease. As you age, the chances of having hypertension increase.[2] This study investigates both diabetes and hypertension because most people with diabetes eventually have hypertension, and hypertension leads to many complications in diabetes.[3]

Several techniques are used to predict diabetes and blood pressure disease to help experts understand them and give a solution to disease diagnosis quickly and make a good decision. They help to predict if the person is diseased or not. Many factors lead to diabetes and blood pressure diseases and might help them to predict if the person diseased or not. They can benefit from the capabilities that Artificial Neural Networks (ANNs) support and its ability in making decisions.

ANN is a mathematical model based on equations or an arithmetic model inspired by the brain's infrastructure and functional aspects of the biological neural networks. ANNs provide many benefits like nonlinearity, input–output mapping, and adaptive, etc. ANNs are widely used in various applications such as business, science, marketing, and medicine. For example, if used in medicine, ANNs can be used to control a lot of health more criteria or it can be used to predict the patient's response to treatment.[4]

The main objective of this research is to benefit from ANNs prediction capabilities. Examine whether an MLP neural network can help to precisely predict if patients are diabetes and/or suffer from blood pressure problems. Also, help determine the factor which has a high influence on these diseases. This study presents a prediction method for both diabetes and blood pressure by using ANNs. Python programming language was used to build the neural network model, test its accuracy, and compare it with other neural networks and classifiers. Two datasets were used to predict the diseases and determine the factors that influence these diseases.

The rest of this paper is organized as follows. Section 2 provides a brief overview of the previous literature and studies related to this topic. Section 3 introduces the proposed methodology and ML algorithms used in this paper. A discussion of the research results is presented in Sec. 4. Our disease diagnosis tool is introduced in Sec. 5. Finally, Sec. 6 provides a conclusion of the research work and directions for future work.

2. Related Work

Nowadays, ANNs are one of the important research areas. They are widely used in science, technology, and medicine with applications in various branches.

Many publications that provide a survey about ANNs and present some advantage of the neural network.[1,4,5] This section discusses some related work that used ANNs in different applications.[2,6]

Maind *et al.*[1] presented a review of ANNs. They discuss many alternative techniques and their several advantages that lead researchers to use it in their work. The authors mention some advantages of the neural networks such as self-organization, flexibility, and learning adaptivity.

Sharma *et al.*[4] provided in their work an overview of ANNs. They concluded that ANN play an essential part in technology in addition to many areas such as medicine, energy, etc. ANNs provide many alternatives and can be applied in multiple applications.

Sabanci *et al.*[2] proposed a computer-based vision method to precisely classifying the wheat grains to bread or durum using ANN to obtain accurate results. The model had trained in 180 wheat grains and tested in 20 wheat grains from a total of 200 grains. The input parameters of the neural network are visual features, twelve main features were determined by images that are taken to a hundred bread and a hundred durum wheat grains. Results show that seven inputs are most accurate to classifying based on the Cfs Subset Eval algorithm to simplify the ANN model. When compared their proposed method with an MAE of 9.8×10^{-6}, the result showed that the proposed method can be easily integrated into the industry to automatically classify agricultural grains.

Bisi *et al.*[6] proposed a neural network-based software reliability model that aims to predict the cumulative number of failures in software programs. They used a feed-forward architecture, the input of the model is encoded using an exponential/logarithmic function, the output of the prediction is the cumulative number of failures, the approach is applied to 18 different datasets. The result of the approach is found acceptable for all different datasets. They concluded that using ANNs is an effective technique to predict software reliability.

Using artificial neural networks in medical prediction helps software engineers and medical experts to examine, diagnose, and make good decisions. There are many reported studies and research works concerning the application of ANNs in medical prediction and diagnosis.[7–10]

Agrawal *et al.*[7] provided a survey about using ANN in medicine, especially medical diagnostics for cancer. They used a neural network to detect and predict cancer cells. The survey showed that many researchers had used ANNs in several areas of medicine and that ANNs provide accurate results. Because of that, the authors used a neural network to detected cancer cells and conclude that neural networks are effective techniques to detect and predict cancer cells. Their study showed a 97.1% accurate result.

Also, Bordoloi and Sarma[8] presented a study of prediction a protein secondary structure by using ANN. They considered four proteins with different functions and used a Multi-Layer Perceptron feedforward model. The backpropagation algorithm is used to update the weight of the network. This work designed three ANN classifiers, which classify the amino acids, protein primary structure, and secondary structure of proteins. The approach has been applied to “amino acids” to predict protein secondary structure from three types of proteins. The results showed 100% accuracy during the training phase while validating the learning phase with the same set.

Gueli *et al.*[9] used ANNs to predict the risk factors that lead to cardiovascular disease based on people’s lifestyle and relate these factors with the probabilities of hit diseases that will threaten their lives to the death. They used a data sample for 276 people aged between 26 and 69 years (both men and women). A sample of 246 of the total data was used in the training phase and the remaining part was used in the testing phase. The input data to the ANN model is (sex, age, build, weight, marital status, number of children, number of cigarettes smoked/day, amount of wine, and cups of coffee), while the output data was (high/low cholesterolemia, HDL cholesterol, triglyceridemic). The results of this study showed that the use of ANNs has a good performance to determine the risk factors.

Zheng *et al.*[10] conducted a study to distinguish the use of ANNs to predict the risk of death in -acute-on-chronic hepatitis B liver failure (ACHBLF) of patients from the use of a scoring system that assess the severity of chronic liver disease (The Model of End-Stage Liver Disease (MELD)). The authors used data from 402 patients with the same routine life and performed the required tests for the liver. The results showed that ANNs produce high performance and accuracy than MELD. Many of them found that neural networks are flexible and provide accuracy in the prediction and diagnosis.

2.1. *Diabetes prediction related work*

This section presents work related specifically to diabetes prediction and diagnosis.[1,11–17]

Luangruangrong *et al.*[11] presented some new factors (such as smoking and alcohol consumption) to improve accuracy in diabetes prediction and develop diabetes prediction tools. The proposed diabetes prediction method is based on BNN. They used a dataset taken from the Bangkok Hospital. The data is for Thai people (2000 cases between the years 2010 and 2012). It includes 1,140 diabetic patients’ cases and 860 healthy person cases. The results of the proposed factors in different experiments introduce a value of accuracy that exceeded the baseline.

Others proposed a user-friendly tool built using MATLAB.[14] The tool has a GUI which acts as a medium between patients and doctors. It helps doctors by presenting the results within a fraction of a second. The study uses a dataset obtained from an Excel sheet, testing 6 different parameters that are used as input to diagnose whether the person is diabetic or not. The used variables are (1) Number of times pregnant, (2) body mass index (BMI), (3) plasma glucose, (4) diastolic blood pressure, (5) triceps skinfold thickness, and (6) diabetic pedigree function. Output is either 0 or 1 (diabetic and normal patients). The training phase is done in several steps. The accuracy obtained of the proposed system was 81% with a smaller number of iterations.

Sandhamet *et al.*[17] used ANNs to predict blood glucose levels (BGL) to control glycemia. The goals of this study are to educate and advise Type1 diabetic patients. The inputs are an insulin regime, anticipated diet, exercise schedule, and blood glucose level. Training phase was performed using backpropagation, and the study was applied to two diabetic patients. The patients were regularly monitored and daily recorded their BGLs, insulin regime, diet, and exercise activity for 10 days. The results of the prediction are very close to the measured values. So, using ANNs could be beneficial to diabetic patients if accurate predictions were assured.

Pappadaet *et al.*[18] proposed a feed-forward neural network model for real-time prediction of glucose in diabetic patients. The inputs are continuous glucose monitoring (CGM), insulin dosages, metered glucose values, nutritional intake, lifestyle, and emotional factors. The proposed model was implemented using the predictive horizon (PH) of 75 min. Their model trained using 17 patients and tested on 10 patients who are not included in the training phase. The neural network model was trained via backpropagation training algorithm. The model predicts 88.6% of normal glucose concentrations, 72.6% of hyperglycemia, and 2.1% of hypoglycemia. Based on the results, the proposed model provides intelligent methods for therapeutic guidance.

Zainuddin *et al.*[19] proposed an ANN that predicts the level of glucose diabetic patients' blood. The proposed system predicted by splitting the day into different intervals: morning (before breakfast, after breakfast), afternoon (before lunch, after lunch), evening (before dinner, after dinner), and night (before sleeping). The patient should answer daily information questionnaire (time of glucose measurement, blood glucose concentration, short-acting insulin injection, long-acting insulin injection, food intake, stress, and exercise). They used a dataset from one patient covering 77 days. The results showed that the proposed system is a powerful and more accurate model for blood glucose prediction when compared with other neural network models, which used the same dataset.

Pappada *et al.*[20] proposed neural network models using Neuro-Solutions software with variable predictive windows of 50–180 min. The proposed models were trained using datasets that include data of 11–17 patients and tested on patient data that are not included in the neural network formulation. They obtained data by filling a diary documenting blood glucose reading, insulin doses, carbohydrate intake, hyperglycemic and hypoglycemic symptoms, and lifestyle. The network was trained using data collected from 18 patients with a constant predictive window of 100 min. The network has been tested using continuous glucose monitoring and diary data from patients who were not included in the training data. Patients used CGM for a period of 3–9 days. Results showed that using such models might lead to better glycemic control.

Katayamaet *et al.*[22] developed a blood pressure measurement device that can measure blood pressure continuously using the Fiber Bragg Grating (FBG) sensor, FBG was used to measure the pulse wave, and they predicted the blood pressure using both PLSR and ANN. They conducted a study using 77 participants (44 males and 33 females). The participants' ages ranged between 21 and 87. 132 data recorded were obtained, 100 are used for training, and 32 for testing. The results showed that the prediction accuracy using PLSR was low. Also, they found that ANNs are better in prediction because the effect of individuals differences was lower.

Asril *et al.*[23] conducted a study to identify factors that affect the healthy lifestyle behaviors of diabetes patients in rural Indonesia. The suggested factors are *severity, barriers, family support, bonding social capital, susceptibility, and chance locus of control.* Also, they introduced a health belief model and tested it along with the suggested lifestyle factors to describe the behaviors.

According to Weizmann Institute of Science,[24] a study reported in Nature Medicine claims that a computer algorithm can help predict the occurrence of gestational diabetes in the early stages of pregnancy, or even before pregnancy.

Table 1. Using ANNs in different applications.

No	Area	Usage / Example(s)	Accuracy	Reported in Refs.
1	Introduction	Introduces ANNs and shows the advantage of using them	—	1, 4
2	Science	Classification of wheat grains	99.9273%	2
3	Technology	Prediction of the cumulative number of failures in the software	—	9
4	Medical	Prediction for cancer	97.1%	6

Table 2. Using ANN for diabetes prediction.

No	Method(s)	Factors	Accuracy	Reported in Refs.
1	Diabetes prediction tools	Smoking alcohol	83.65%	13
2	User-friendly software tool	Times pregnant, Body mass index, plasma glucose, diastolic blood. Pressure, triceps skinfolds thickness, diabetic pedigree function	81%	14
3	Predicting blood glucose levels	Insulin regime, anticipated diet, exercises schedule, blood glucose level	—	17
4	Real-time prediction of glucose	CGM, insulin dosages, metered glucose values, nutritional intake, lifestyle, and emotional factors	88.6%	18
5	Predicting the level of glucose in the blood	Insulin injection, food intake, stress, and exercise	—	19
6	Diabetes	Blood glucose readings, insulin doses, lifestyle carbohydrate intake, hyperglycemic, hypoglycemic symptoms	—	20

The study analyzed data of 600,000 pregnancies available from Clalit Health Services. They found that gestational diabetes may be predicted and prevented using nutritional and lifestyle changes.

Previous studies used ANNs for diabetes and blood pressure prediction, some of them shown precise results. Tables 1 and 2 provide a summary of related studies.

Table 2 presents studies and papers that used ANNs for diabetes prediction.

3. Methodology

This section presents our approach of using neural networks to predict diabetes and blood pressure diseases. This section briefly explains the process phases.

- **Factors selection**: suitable input factors are selected carefully.
- **Dataset building**: two online datasets were extracted.
- **Preprocessing**: the dataset was preprocessed before evaluation. Also, the dataset was checked for outliers and extreme values (dataset cleaning).
- **Training and verification**: the neural network architecture was defined and the training process was performed.
- **Finally**: applying a neural network and use it in practice to predict the diagnosis (the probability of being diseased or not). This work used a Multi-Layer Perceptron Neural Network architecture and compared the performance measures with two other neural networks (*Support Vector Machine* (*SVM*) and *K-Nearest Neighbors* (*KNN*)).

Section 4 presents the experiment that was conducted to predict both of the blood pressure and diabetes and the found results.

3.1. *Data selection*

This study used two online datasets for Indian patients. The first dataset for diabetes patients and the second one for blood pressure patients. After that, the datasets passed through a filtering process to make them fit the purpose of this study.

The diabetes dataset consists of 768 instances (diagnosed as a diabetic = 500, diagnosed un-diabetic = 268). Also, the blood pressure disease dataset consists of 768 instances (diagnosed as hypertension = 500, diagnosed as not hypertension = 268).

After filtering and removing anomalies (both outliers and extreme values). The diabetes dataset consists of 757 instances (diabetic = 496, un-diabetic = 261). The blood pressure dataset consists of 760 instances (hypertension = 500, not hypertension = 260) (see Table 3).

3.2. *Factors selection*

The diabetes dataset consists of 6 features for each data instance. The input features are *Age, Weight, Fat Ratio, Insulin, and Glucose*, and the output feature is *Diseased*.

Table 3. Datasets instances information.

	Original dataset			After filtering				
	Instances (#)	Diabetic	Un-diabetic	Instances (#)	Diabetic	Un-diabetic	Outliers	Extreme values
Diabetes dataset	768	500	268	757	496	261	11	0
Blood pressure dataset	768	500	268	760	500	260	0	8

To classify the 757 data instances, the output (Diseased) feature is divided into two parts; **0 (diabetic)** and **1 (un-diabetic)**.

Furthermore, the blood pressure disease dataset consists of seven features for each data sample. The input features are *Age, Weight, Fat Ratio, Blood Pressure, Alcohol, and Smoking*, and the output feature is *Diseased*. To classify the 760 data instances, the output (Diseased) feature is divided into two parts; **0 (hypertension)** and **1 (un-hypertension)**. Also, both Alcohol and Smoking factors are classified as **0 (no)** and **1 (yes)**.

3.3. *Multi-Layer Perceptron Neural Network (MLP)*

3.3.1. *Diagnosis of diabetes using MLP*

- **Step 1.** Defining the MLP neural network architecture for the diabetes diagnosis. In this research, we implemented the MLP neural network with five hidden layers feed-forward using Python to predict diabetes diagnosis. The network consists of an input layer, hidden layers, and an output layer against a set of units or nodes. The input layer passes information from 757 instances, and the output layer includes 2 units. Number of input parameters equals the number of input layer units. In the hidden layers, a "Hyperbolic tangent" activation function is used,[10] while in the output layer, a "Softmax" activation function is used.[10,21]
- **Step 2.** Training and testing the neural network using Python. The datasets are considered as input parameters for training and testing the ANN model; out of 757 patient's data (80%) were used for training and the remaining (20%) were used for testing. MLP predicted that 95 out of 152 patients are diabetic when there were 119 patients are diabetics.
- **Result.** Applying MLP on the diabetes dataset. Regarding the *diabetic* diagnosis, the performance metrics values are ($F1$-*score* = 0.84, *Recall* = 0.95, *and Precision* = 0.76). Regarding the *un-diabetic* diagnosis, the metrics values are (*F1-score* = 0.62, *Recall* = 0.49, *and Precision* = 0.85). The model has an accuracy of (77.63%).

3.3.2. *Diagnosis of blood pressure using MLP*

- **Step 1.** Defining the MLP neural network architecture for the pressure diagnosis. A model was built using Python for the MLP neural network with five hidden layers feed-forward to predict diabetes diagnosis. The network consists of an input layer, hidden layers, and an output layer against a set of units or nodes. The input layer passes information from 760 instances, and the output layer includes 2 units. Number of input parameters equals the number of input layer units. In the hidden layers, a "Hyperbolic tangent" activation function is used,[10] while in the output layer, a "Softmax" activation function is used.[10,21]
- **Step 2.** Training and testing the neural network using Python. The datasets are considered as input parameters for training and testing the ANN model; out of 760 patient's data (80%) were used for training and the remaining (20%) were used for testing. MLP predicted that 132 out of 192 patients are diabetic when there were 140 patients are diabetics.
- **Result.** Applying MLP on the blood pressure dataset. Regarding the *diabetic* diagnosis, the performance metrics values are ($F1$-*score* = 0.78, *Recall* = 0.80, *and Precision* = 0.76). Regarding the *un-diabetic* diagnosis, the metrics values are ($F1$-*score* = 0.46, *Recall* = 0.43, *and Precision* = 0.50). The model has an accuracy of (68.75%).

3.4. *K-Nearest Neighbors* (*KNN*)

This section presents the results when KNN was applied to both of the two datasets. Table 4 shows the performance results of the KNN model.

- First, applying the model on the diabetes dataset. Regarding the *diabetic* diagnosis, the metrics values are ($F1$-*score* = 0.79, *Recall* = 0.87, *and Precision* = 0.73). Regarding the *un-diabetic* diagnosis, the metrics values are ($F1$-*score* = 0.59, *Recall* = 0.50, *and Precision* = 0.71).
- Second, applying the model on the blood pressure dataset. Regarding the *hypertension* diagnosis, the metrics values are ($F1$-*score* = 0.69, *Recall* = 0.76, *and*

Table 4. Performance results from applying KNN.

Diagnosis	Precision	Recall	F1-score
diabetic	0.73	0.87	0.79
un-diabetic	0.71	0.50	0.59
hypertension	0.63	0.76	0.69
un-hypertension	0.42	0.29	0.34

Table 5. Performance results from applying SVM.

Diagnosis	Precision	Recall	F1-score
diabetic	0.75	0.91	0.82
un-diabetic	0.71	0.42	0.53
hypertension	0.60	1.00	0.75
un-hypertension	1.00	0.03	0.06

Precision $= 0.63$). Regarding the *un-hypertension* diagnosis, the metrics values are ($F1$-*score* $= 0.34$, *Recall* $= 0.29$, *and Precision* $= 0.42$).

3.5. *Support Vector Machine (SVM)*

SVM was applied to both of the two datasets. This section presents the performance results of the SVM model (see Table 5).

- First, applying the model on the diabetes dataset. Regarding the *diabetic* diagnosis, the metrics values are ($F1$-*score* $= 0.82$, *Recall* $= 0.91$, *and Precision* $=$ 0.75). Regarding the *un-diabetic* diagnosis, the metrics values are ($F1$-*score* $=$ 0.53, *Recall* $= 0.42$, *and Precision* $= 0.71$).
- Second, applying the model on the blood pressure dataset. Regarding the *hypertension* diagnosis, the metrics values are ($F1$-*score* $= 0.75$, *Recall* $= 1.00$, *and Precision* $= 0.60$). Regarding the *un-hypertension* diagnosis, the metrics values are ($F1$-*score* $= 0.06$, *Recall* $= 0.03$, *and Precision* $= 1.00$).

4. Discussion of Results

This section presents a comparison between the performance metrics of KNN, SVM, and MLP (see Table 6).

Table 6. Performance results from applying KNN, SVM, and MLP.

Diabetes Info.			
	KNN	SVM	MLP
F1-score	0.71	0.72	0.76
Recall	0.72	0.74	0.78
Precision	0.72	0.74	0.79
MCC	0.403	0.392	0.515
Accuracy	0.723	0.743	0.776
Pressure Info.			
	KNN	SVM	MLP
F1-score	0.56	0.47	0.68
Recall	0.58	0.60	0.69
Precision	0.55	0.76	0.68
MCC	0.051	0.137	0.246
Accuracy	0.57	0.60	0.68

Table 7. AUCs obtained from applying KNN, SVM, and MLP.

AUC-Diabetes		
KNN	SVM	MLP
0.684	0.666	0.719
AUC-Pressure		
KNN	SVM	MLP
0.523	0.515	0.618

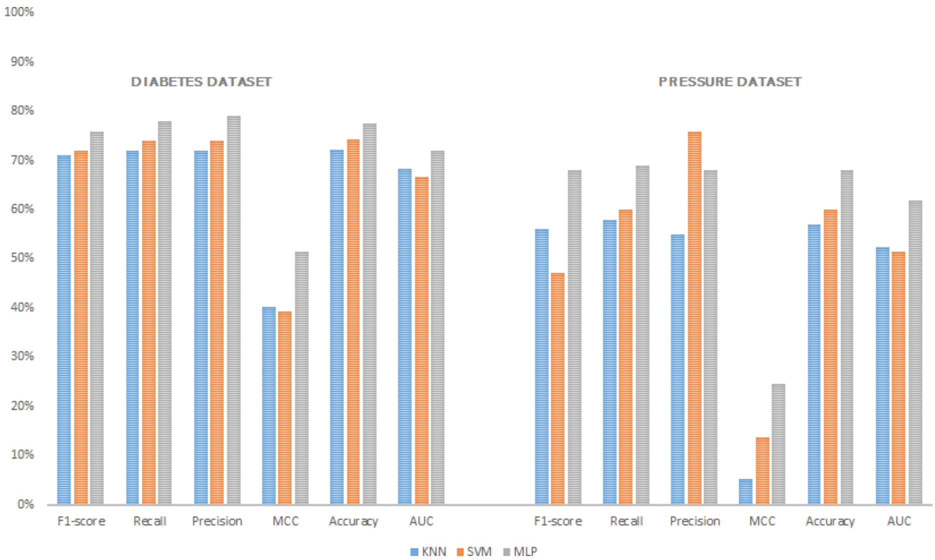

Fig. 1. Performance measures — LSTM, SVM, and KNN.

Figure 1 illustrates the performance differences between the three ML algorithms.

The results of our experiments indicate that MLP Neural Network correctly predicts and diagnosis the probability of being diseased or not, and the performance can be significantly increased compared with both SVM and KNN (see Fig. 1).

Based on the found metrics (see Table 6 and Fig. 1), we make the following observations:

- **Diabetes:**
  - F1-score results show a **4%** improvement for MLP compared with SVM. Also, it shows a **5%** improvement for MLP compared with KNN.
  - MCC values improved by **12.3%** compared to SVM and by **11.2%** compared to KNN; values show that MLP outperforms the other algorithms in detecting and diagnosing the diabetes disease.

- **Blood Pressure:**
  - F1-score results show a **21%** improvement for MLP compared with SVM. Also, it shows a **12%** improvement for MLP compared with KNN.
  - MCC values improved by **10.9%** compared to SVM and by **19.5%** compared to KNN, which show that MLP outperforms the other algorithms in detecting and diagnosing the blood pressure disease.

  Compared with KNN and SVM, it has been found that MLP achieves a slight performance improvement in diagnosing both of diabetes and blood pressure diseased.

4.1. *The ability of models — MLP, KNN, and SVM*

Also, the Receiver Operating Characteristic (ROC)[25] curves present a visual perception of the specificity and sensitivity for all possible cut-offs. Figures 2 and 3 summarize the areas under ROC curves obtained from training and testing the three ML algorithm on both of the two diseases.

Figures 2 and 3 show the area under the ROC curves (AUCs) which indicate that MLP AUCs values are better compared with SVM and KNN. Diabetes prediction (MLP = 0.719, SVM = 0.666, and KNN = 0.684), and blood Pressure (MLP = 0.618, SVM = 0.515, and KNN = 0.523). Thus, a well-trained MLP model could successfully identify both of the two diseases, with high accuracy and better AUCs (see Table 7).

Figures 4 and 5 show the predicted pseudo probabilities.[26] The graphs display clustered box-plots of predicted pseudo-probabilities of the correctly classified cases

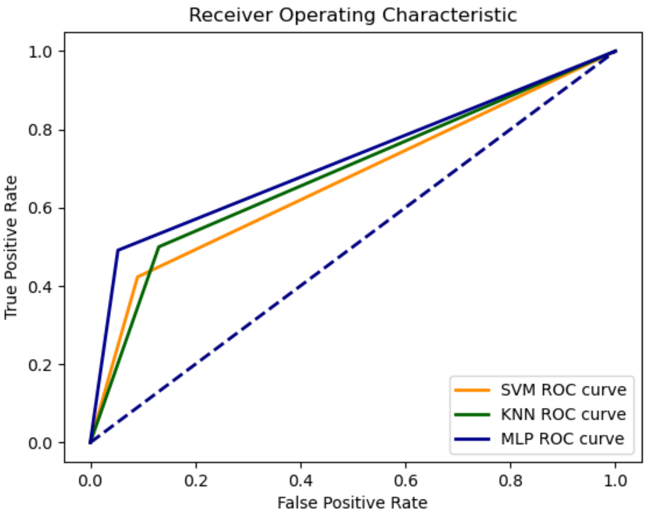

Fig. 2. ROC curve for diabetes.

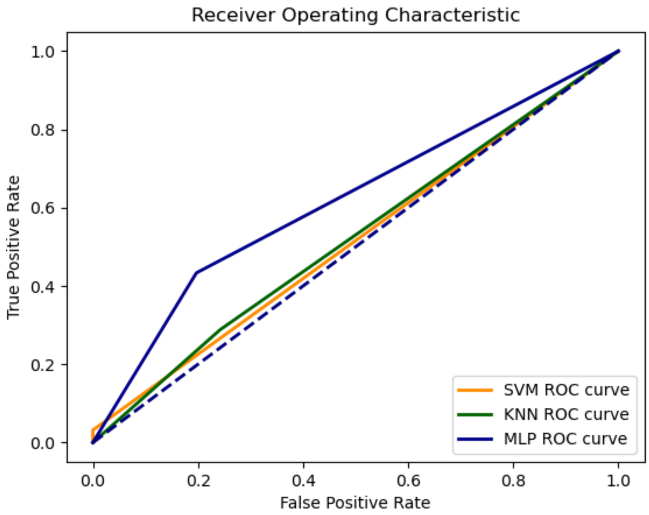

Fig. 3. ROC curve for pressure disease.

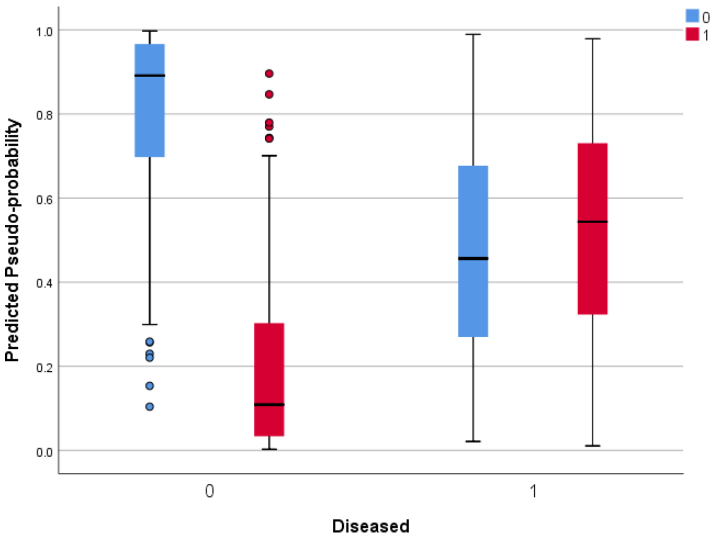

Fig. 4. Predicted pseudo-probability for diabetes.

of categories "zero" and "one". Also, they show some incorrect classified cases of both categories that were observed below the cut point 0.5.

Figure 6 shows the independent variables that are most important for predicting diabetes, where "glucose" is the most significant variable and "weight" is the less significant variable. Figure 7 presents the independent variable that is most

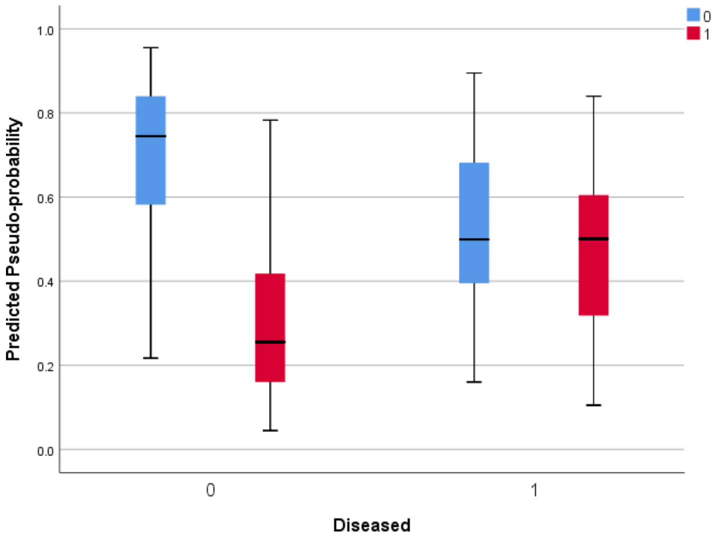

Fig. 5. Predicted pseudo-probability for blood pressure.

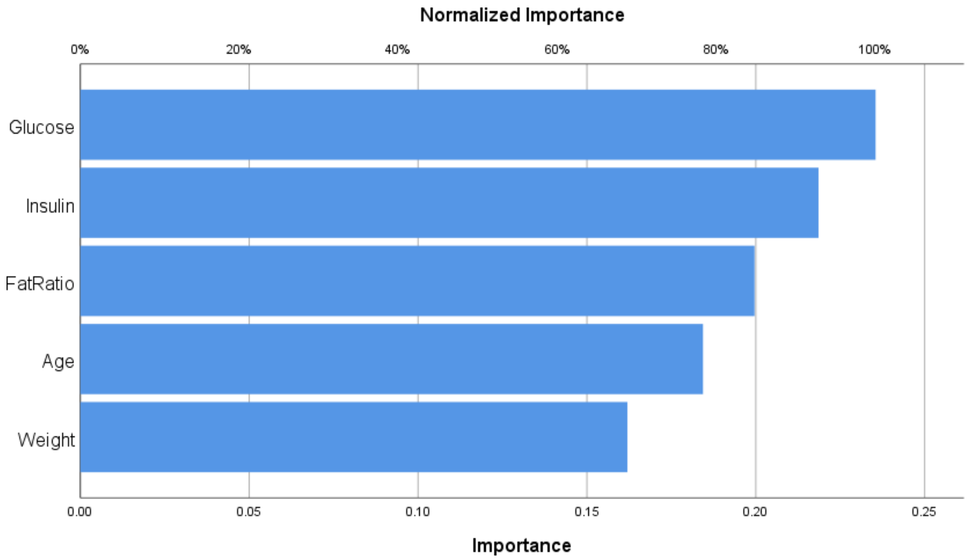

Fig. 6. Diabetes independent variables.

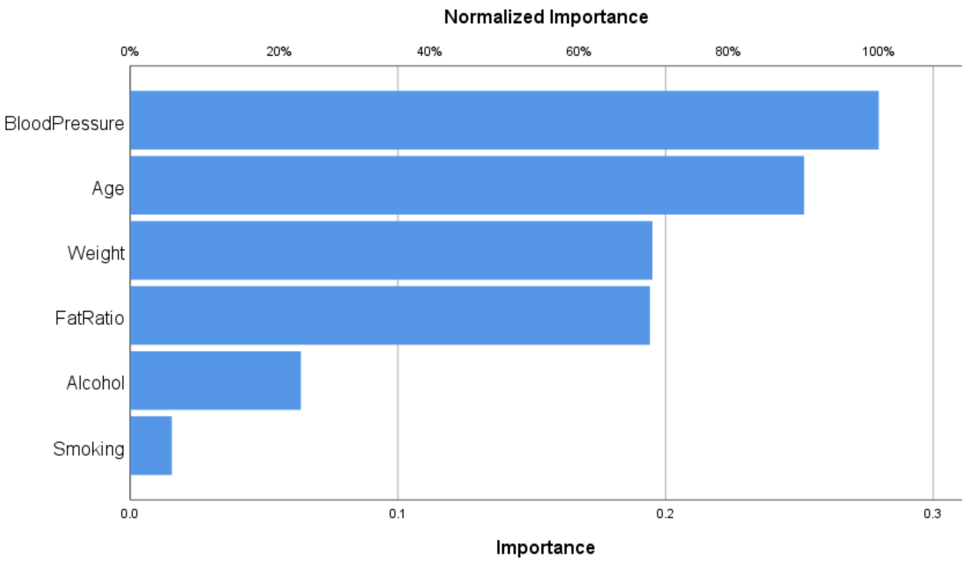

Fig. 7. Pressure independent variables.

important for predicting pressure disease, where “blood pressure” is the most significant variable and “smoking” is the less significant variable.

5. D&Ppredict — A Disease Diagnosis Tool

Diseases such as diabetes and blood pressure are considered as death causes. A tool called **D&Ppredict** (see Fig. 8) has been introduced to help expertise diagnose diseases faster and make a good decision to treat the patients. This tool helps experts predict if the person is diseased or not. It has been developed using Python and can be installed on any Operating System. The tool uses XML files that are generated using SPSS.

The prediction is based on many factors. The selected factors are *Age, Weight, Fat Ratio, Glucose, Insulin, Blood Pressure, Smoking,* and *Alcohol*. *Age, Weight,* and *Fat Ratio* are common factors between these two diseases. The tool helps predict if a person is diseased or not.

D&Ppredict allows the user to enter some factors (each factor has a weight/value in the XML file). The XML file contains an ID for all values of the factors. Using IDs the tool returns a weight for the values entered by the user. If the XML file does not have a weight/value for any factor, the tool generates a random number (between -1 and 1) as a weight for this factor. After that, the tool calculates the “weighted sum” of the input factors values and adds a bias as a third input (as shown in Eq. (5.1)).

$$Y = \sum(\text{weight} * \text{input}) + \text{bias}. \tag{5.1}$$

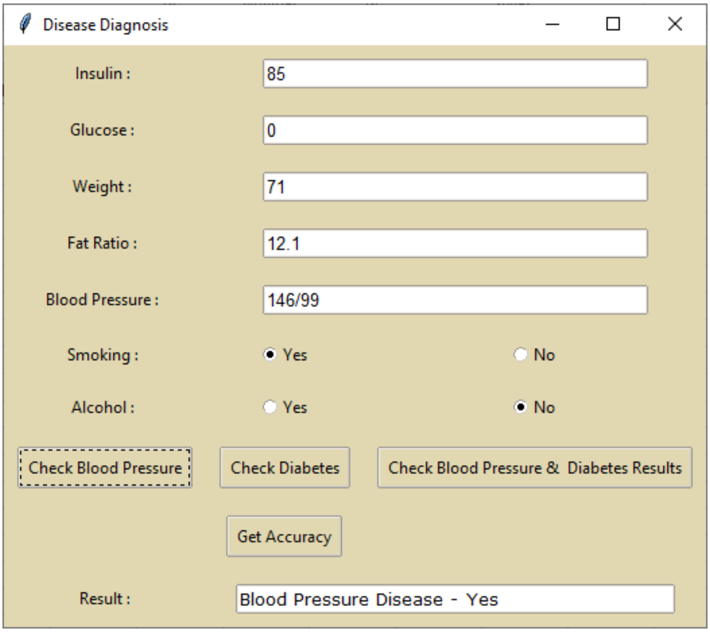

Fig. 8. Disease diagnostic tool (D&Ppredict).

Then apply the activation and determine whether the weighted sum is greater than a threshold value (threshold is equivalent to bias) (see Eq. (5.2)).

$$F(x) = \tanh(x) = (2/(1 + e^{-2x})) - 1. \tag{5.2}$$

The result of the second equation will be between 0 and 1. If the result approximates 0, then the person is "Not Infected" with this disease. If the result approximates 1, then the person is "Infected" with this disease.

The tool provides accurate results. It has been tested using random data from a real dataset and compared with the results generated by the tool. Based on the comparison, the accuracy of the result is more than 90%.

6. Conclusions and Future Work

In recent years, several researchers use ANNs to predict, classify, diagnose, help surgeons and physicians, and the community. This research applied ANNs to predict diabetes and hypertension with reasonable accuracy.

Two online datasets for Indian patients were used, which contains 768 records each. The first for diabetes and the second for hypertension. They were divided into training and test cases and applied dataset variations (20% test and 80% training).

The performance of Multi-Layer Perceptron (MLP) is compared with two other algorithms (SVM and KNN) performances. The results present that MLP diagnosis the diseases more accurately.

The results indicating 77.6% accuracy in diabetes and 68% in hypertension. Also, F1-score and MCC values show improvement for MLP compared with both of the two algorithms. This shows that the used method is effective and improves disease prediction.

Further prospective studies are required to confirm our findings. Obvious directions for future work include using different datasets. Other variables, in addition to the variables used in this work, will be included in the process of identifying the diseases. Also, the same process might be repeated using other machine learning techniques and algorithms.

References

1. Maind S., Wankar P., Research paper on basic of artificial neural network, *Int. J. Recent Innovation Trends Comput. Commun.* **2**(1):96–100, 2014.
2. Sabanci K., Kayabasi A., Toktas A., Computer vision–based method for classification of wheat grains using artificial neural network, *J. Sci. Food Agricult.* **97**(8):2588–2593, 2017.
3. Diabetes and High Blood Pressure: What is the relationship?, available at https://diabetesstrong.com/diabetes-and-high-blood-pressure/. Last Accessed July 15, 2020.
4. Sharma K., Naik A., Patel P., Study of artificial neural networks, *Int. J. Adv. Res. Trends Eng. Technol.* (*IJARTET*) **2**(4):46–48, 2015.
5. Amato F. *et al.*, Artificial neural networks in medical diagnosis, *J. Appl. Biomed.* **11**(2):47–58, 2013, ISSN: 1214-021X.
6. Agrawal S., Agrawal J., Neural network techniques for cancer prediction: A survey, *Procedia Comput. Sci.* **60**:769–774, 2015.
7. Bordoloi H., Sarma K. K., Protein Structure prediction using multiple artificial neural network classifier, *Soft Computing Techniques in Vision Science*, Studies in Computational Intelligence, Vol. 395, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-25507-6_12.
8. Gueli N. *et al.*, The influence of lifestyle on cardiovascular risk factors: Analysis using a neural network, *Archiv. Gerontol. Geriatrics* **40**(2):157–172, 2005.
9. Bisi M., Goyal N., Software reliability prediction using neural network with encoded input, *Int. J. Comput. Appl.* **47**(22):46–52, 2012.
10. Glorot X., Bengio Y., Understanding the difficulty of training deep feedforward neural networks, *Proc. Thirteenth Int. Conf. Artificial Intelligence and Statistics* (*AISTATS*), Chia Laguna Resort, Sardinia, Italy, Vol. 9, pp. 249–256, 2010.
11. Zheng M. H. *et al.*, A model to predict 3-month mortality risk of acute-on-chronic hepatitis *b* liver failure using artificial neural networks, *J. Viral Hepatitis* **20**(4):248–255, 2013.
12. Kohli S., Miglani S., Rapariya R., Basic artificial neural network, *Int. J. Comput. Sci. Mobile Comput.* **3**(9):745–751, 2014.
13. Luangruangrong W., Rodtook A., Chimmanee S., Study of type 2 diabetes risk factors using neural network for thai people and tuning neural network parameters, *2012 IEEE Int. Conf. Systems, Man, and Cybernetics* (*SMC*), Seoul, pp. 991–996, 2012, doi: 10.1109/ICSMC.2012.6377858.

14. Joshi S., Borse M., Detection and prediction of diabetes mellitus using back-propagation neural network, *2016 Int. Conf. Micro-Electronics and Telecommunication Engineering* (*ICMETE*), Ghaziabad, pp. 110–113, 2016, doi:10.1109/ICMETE.2016.11.
15. Mohapatra S. K., Mihir J. K. S., Mohanty N., Detection of diabetes using multilayer perceptron, *Int. Conf. Intelligent Computing and Applications, Advances in Intelligent Systems and Computing*, Bhaskar M. A., Dash S. S., Das S., Panigrahi B. K. (eds.), Springer, Singapore, 2019.
16. Mishra S., Tripathy H. K., Mallick P. K., Bhoi A. K., Barsocchi P., EAGA-MLP an enhanced and adaptive hybrid classification model for diabetes diagnosis, *Sensors* **20**(14):4036, 2020.
17. Sandham W. *et al.*, Blood glucose prediction for diabetes therapy using a recurrent artificial neural network, *9th European Signal Processing Conf.* (*EUSIPCO 1998*), Rhodes, pp. 1–4, 1998.
18. Pappada S. M.*et al.* Neural network-based real-time prediction of glucose in patients with insulin-dependent diabetes, *Diabetes Technol. Therapeutics* **13**(2):135–141, 2011.
19. Zainuddin Z., Ong P., Ardil C., A neural network approach in predicting the blood glucose level for diabetic patients, *Int. J. Comput. Intell.* **5**(1):72–79, 2009.
20. Pappada S. M., Cameron B. D., Rosman P. M., Development of a neural network for prediction of glucose concentration in type 1 diabetes patients, *J. Diabetes Sci. Technol.* **2**(5):792–801, 2008, https://doi.org/10.1177/193229680800200507.
21. Szegedy C. *et al.*, Intriguing properties of neural networks, available at https://arxiv.org/abs/1312.6199v4.
22. Katayama K. *et al.*, Improvement of blood pressure prediction using artificial neural network, *2018 IEEE Int. Symp. on Medical Measurements and Applications* (*MeMeA*), Rome, pp. 1–5, 2018, doi:10.1109/MeMeA.2018.8438657.
23. Asril, N. M., Tabuchi, K., Tsunematsu, M., Kobayashi, T., Kakehashi, M., Predicting healthy lifestyle behaviours among patients with type 2 diabetes in rural bali, Indonesia. *Clinical Medicine Insights*, *Endocrinology and Diabetes*, Vol. 13, 2020, https://doi.org/10.1177/1179551420915856.
24. Weizmann Institute of Science, Algorithm can predict gestational diabetes using nutritional and lifestyle changes, available at https://medicalxpress.com/news/2020-01-algorithm-gestational-diabetes-nutritional-lifestyle.html. Last Accessed July 04, 2020.
25. Woods K., Bowyer K. W., Generating ROC curves for artificial neural networks, *IEEE Trans. Med. Imag.* **16**(3):329–337, 1997, doi:10.1109/42.585767.
26. Bevilacqua V. *et al.*, A face recognition system based on pseudo 2d hmm applied to neural network coefficients, *Soft Comput.* **12**(7):615–621, 2008.